2009

The Effect of Heritability Estimates on High-Density Single Nucleotide Polymorphism Analyses with Related Animals

Abebe T. Hassen
Iowa State University

Santiago Avendano
Aviagen Ltd.

W. G. Hill
University of Edinburgh

Rohan L. Fernando
Iowa State University, rohan@iastate.edu

Susan J. Lamont
Iowa State University, sjlamont@iastate.edu

Published December 5, 2014

The effect of heritability estimates on high-density single nucleotide polymorphism analyses with related animals[1]

A. Hassen,* S. Avendano,† W. G. Hill,‡ R. L. Fernando,* S. J. Lamont,* and J. C. M. Dekkers*[2]

*Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames 50011;
†Aviagen Ltd., Newbridge, Midlothian, EH28 8SZ, United Kingdom;
and ‡Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JR, United Kingdom

ABSTRACT: Analysis of high-density SNP data in outbred populations to identify SNP that are associated with a quantitative trait requires efficient ways to handle large volumes of data and analyses. When using mixed animal models to account for polygenic effects and relationships, genetic parameters are not known with certainty, but must be chosen to ensure proper evaluation of SNP across chromosomes and lines or breeds. The objectives of this study were to evaluate the influence of heritability on the estimates and significance of SNP effects, to develop efficient computational strategies for analysis of high-density SNP data with uncertain heritability estimates, and to develop strategies to combine SNP test results across lines or breeds. Data included sire SNP genotypes and mean progeny performance from 2 commercial broiler breeding lines. Association analyses were done by fitting each SNP separately as a fixed effect in an animal model, using a range of heritabilities. The heritability used had a limited impact on SNP effect estimates, but affected the SE of estimates and levels of significance. The shape of the frequency distribution of *P*-values for the test of SNP effects changed from a highly skewed L-shaped curve at low heritability to a right-skewed distribution at high heritability. The *P*-values for alternative heritabilities could, however, be derived without reanalysis based on a strong linear relationship ($R^2 = 0.99$) between differences in log-likelihood values of models with and without the SNP at different levels of heritabilities. With uncertain estimates of heritability, line-specific heritabilities that ensure proper evaluation of SNP effects across lines were determined by analysis of simulated sire genotypes and by permutation tests. Resulting heritability estimates were between those obtained from the entire breeding populations and those obtained from the data included in the sample data set. In conclusion, the uncertainty of heritability estimates has a limited impact on SNP effect estimates in association analyses, but a large impact on significance tests. The impact of heritability on tests can, however, be dealt with in a computationally efficient manner by using the strong linear relationship between model statistics under alternate levels of heritability. These approaches allow efficient analysis of large numbers of SNP for multiple traits and populations and pooling of results across populations.

Key words: association test, high-density single nucleotide polymorphism, linkage disequilibrium mapping

J. Anim. Sci. 2009. 87:868–875
doi:10.2527/jas.2008-0852

INTRODUCTION

Although selection programs on BLUP EBV have sustained rapid genetic progress for some traits, other traits have low heritability or are difficult or expensive to measure. For such traits, markers associated with QTL can be used to bring about genetic change through marker-assisted selection (Meuwissen and Goddard,

[1]The authors acknowledge contributions from Jim McKay, John Ralph, Stewart Brown, and Alfons Koerhuis to this project. Financial support and data were provided by Aviagen Ltd. The SNP assays were developed with the support of the USDA-ARS and the USDA-Cooperative State Research, Education, and Extension Service (CSREES) National Research Initiative Competitive Grants Program (NRICGP), and through the efforts of Hans Cheng (Avian Disease and Oncology Lab, ARS, East Lansing, MI), William Muir (Purdue University, West Lafayette, IN), Gane Wong (University of Alberta, Canada, and Beijing Institute of Genomics, China), Martien Groenen (Wageningen Agricultural University, the Netherlands), and Huanmin Zhang (Avian Disease and Oncology Lab, ARS) for their work on USDA-CSREES-NRICGP proposal 2004-05434 titled "Validation and characterization of a high-density chicken SNP map." This project was also supported by the Iowa Agriculture and Home Economics Experiment Station (Ames; Project No. 3600) and by Hatch Act and State of Iowa Funds.

[2]Corresponding author: jdekkers@iastate.edu

Received January 8, 2008.
Accepted November 7, 2008.

1996). Genotypes from high-density SNP panels can now be used to scan the genome for QTL by linkage disequilibrium (**LD**) or association analyses. A review of statistical procedures for LD mapping of QTL can be found in Dekkers et al. (2006). Although methods are now available to fit all SNP for breeding value estimation simultaneously (Meuwissen et al., 2001), the properties of these methods for QTL mapping are not well understood, and most association analyses are based on fitting SNP separately or as haplotypes (Dekkers et al., 2006). When individuals are related, these analyses must use animal models with relationships to avoid bias of SNP effects and significance tests (Kennedy et al., 1992), but errors in estimates of heritability may influence results (Kennedy, 1991) and valid estimates may not be available if genotyped individuals are a selected sample from the population. In view of their large computational requirements, efficient methods for analysis and reanalysis (for different heritabilities) of high-density SNP data and for combining results across multiple populations are needed. Thus, the objectives here were to 1) evaluate the effect of heritability estimates on association analyses of high-density SNP data in a progeny mean design, 2) develop efficient computational strategies for analysis of high-density SNP data with uncertain heritability estimates, and (3) develop efficient computational strategies to combine association results across populations. This work represents results from methodological work conducted to facilitate LD mapping in multiple commercial broiler breeding lines.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data used were part of already existing data that are collected through the routine breeding program of Aviagen Ltd.

Source of Data

Data were provided by Aviagen Ltd. and included SNP genotypes, mean progeny performance, and pedigree information for 158 and 175 sires from 2 commercial breeding broiler lines (lines I and II) that were used for breeding within these lines during a given time period. The mean number of progeny per sire was 319 and ranged from 2 to 948.

Phenotypic data available for each sire included mean progeny BW at 40 d of age adjusted for fixed effects of sex, age, hatch, and mating group, and for random effects of one-half the EBV of the mate and for estimates of common environmental effects of full sibs. Sires were genotyped for 6,000 SNP across the genome based on an Illumina DNA test panel (Illumina, San Diego, CA). Initial SNP assay development for this panel was coordinated by H. Cheng (USDA-ARS, Avian Disease and Oncology Laboratory, East Lansing, MI), which resulted in a 3K SNP panel with genome-wide coverage, with SNP chosen from those identified by an SNP discovery consortium (Wong et al., 2004). A file titled "Database of SNP used in the Illumina Corp. Chicken Genotyping Project" that describes the original 3K SNP panel is accessible at http://poultry.mph.msu.edu/resources/resources.htm#SNPs (last accessed Jan. 13, 2009). To complement the 3K panel, another 3K SNP across the genome were chosen from the Consortium SNP results to fill in gaps and to increase the density in some regions. Because results were not expected to differ between chromosomes, for the purpose of this study, SNP on chromosomes 1 (n = 959) and 4 (n = 398) were used. The extent of LD measured by r^2 in the populations analyzed extended over relatively short distances, with 24 and 10% of markers within 0.5 cM having LD greater than 0.50 and 0.80. Further details on LD in these populations can be found in Andreescu et al. (2007).

Data Analysis

At each SNP locus, sire genotypes were assigned values of 0, 1, or 2 based on the number of copies of the 0 allele they carried. These values were then included as covariates in a mixed model analysis to estimate the allele substitution effect for each SNP. These analyses were conducted separately for each SNP and for each line. Sires used in the study were from a commercial population that is under selection. Evaluation of the pedigree structure showed that sires within each line belonged to several groups of half-sib families; hence, mixed models that account for relationships among sires were used. The basic single-SNP model used to evaluate the association at each locus was

$$\mathbf{Y} = \mu 1 + \mathbf{Xb} + \mathbf{Zs} + \mathbf{e},$$

where $\mathbf{Y}$ is the n × 1 vector of adjusted progeny means for n sires; μ is the intercept; $\mathbf{b}$ is the fixed SNP allele substitution effect; $\mathbf{s}$ is the n × 1 vector of random sire polygenic effects; $\mathbf{e}$ is the n × 1 vector of random residuals; $\mathbf{X}$ is the vector of number of copies of the 0 allele carried by each sire at the SNP; and $\mathbf{Z}$ is the incidence matrix relating random sire polygenic effects to $\mathbf{Y}$.

The following model expectations and variances were assumed:

$$E\begin{bmatrix} Y \\ s \\ e \end{bmatrix} = \begin{bmatrix} \mu 1 + X\beta \\ 0 \\ 0 \end{bmatrix}, \text{ and}$$

$$V\begin{bmatrix} \mathbf{s} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_s^2 & 0 \\ & \sigma_e^2\mathbf{D}^{-1} \end{bmatrix},$$

where σ_s^2 is the sire variance, σ_e^2 is the residual variance, $\mathbf{D}$ is the n × n diagonal matrix with the number of progeny for each sire on the diagonal to provide the

appropriate residual variance for each progeny mean, and **A** is the additive genetic relationship matrix. The residual variance can be expressed as a function of heritability (h^2) as follows:

$$\sigma_e^2 = \left(1 - h^2/4\right)\sigma_p^2,$$

where σ_p^2 is the phenotypic variance and

$$\sigma_e^2 / \sigma_s^2 = \left[\left(4 - h^2\right) / h^2 \right].$$

Mixed models were fitted for a given level of heritability by using PROC MIXED (SAS Inst. Inc., Cary, NC). The additive genetic relationship matrix was based on 4 generations of pedigree and was calculated by using the SAS macro LORG developed by Zhang (2005). Tests of significance of SNP effects were based on a likelihood ratio test (Searle, 1987) using the log of the model likelihood (including fixed and random effects) for the full and reduced models, where the reduced model included all effects except the allele substitution effect but used the same value of heritability as the full model, under the assumption that variance explained by an individual SNP is expected to be small and does not affect polygenic variance. Significance levels were obtained from a χ^2 distribution with 1 df.

To evaluate the impact of heritability on results, models were fitted by using different levels of heritability. Estimates of heritability obtained in routine genetic evaluation by Aviagen Ltd. using multigeneration data from the entire breeding population were 0.23 for line I and 0.31 for line II. Results of association analyses based on these values were used as a base to evaluate the impact of alternate heritabilities, which were created by increasing or decreasing the sire variance in each line by up to 50% without changing the total phenotypic variance. The 11 resulting heritability values ranged from 0.11 to 0.34 for line I and from 0.12 to 0.47 for line II. For each SNP on chromosome 1, association analyses were carried out for each line separately based on each level of heritability to evaluate the impact of heritability on the estimate and test of SNP effects and to develop an empirical relationship of model log-likelihoods for alternate levels of heritability. The purpose of the latter relationship was to enable rapid computation of SNP test results for alternative levels of heritability without requiring complete reanalysis. The same analyses were also conducted by using SNP on chromosome 4 to validate the relationships that were derived based on chromosome 1 SNP.

RESULTS AND DISCUSSION

Descriptive Statistics

Of the SNP genotyped on chromosomes 1 and 4, 24% were not segregating in line I and 28% were not segregating in line II. The mean minor allele frequency among segregating loci was 0.23 (SD = 0.15). Scatter plots of adjusted progeny means (not shown) revealed 2 sires with extremely low progeny means, but these had only 2 progeny each and were not excluded from the analysis because they have a minor influence on the results.

Nearly 77% of the sires had at least 1 half-sib with progeny data. These sires were from 82 half-sib families, with sizes ranging from 2 to 9, of which 23 families had at least 5 members. The mean additive relationship among sires within a line was 6%.

SNP Effect Solutions and SE

The influence of heritability on SNP effect solutions and their SE is shown in Table 1. Slopes for the regression of SNP effect solutions based on alternate heritabilities on corresponding estimates from analyses that used base heritabilities ranged from 1.02 for low alternate heritabilities to 0.98 for high alternate heritabilities. Thus, there was limited evidence of bias in SNP effect estimates from using different values for heritability. This is consistent with statistical theory, which shows that, under normality, estimates of fixed effects are not biased by the use of an incorrect variance-covariance matrix for residuals (Searle, 1987). Kackar and Harville (1981) provided a detailed proof of unbiasedness of estimates of fixed effects from 2-stage estimation and prediction procedures, in which variances estimated from the data were used to solve for fixed and random effects. Sorensen and Kennedy (1986) also showed that when true variances were replaced by values estimated from data, estimates of genetic and environmental effects were unbiased.

The heritability used for analysis did, however, have a large impact on the SE of SNP effects. As illustrated in Table 1, the mean SE of SNP effect estimates increased with heritability, as did the coefficients of regression of SE for alternate levels of heritability on those for base heritabilities. These results are consistent with those of Littell et al. (1998, 2000), who noted no apparent changes of fixed effect solutions and linear contrasts for different variance-covariance structures, but emphasized the need for careful evaluation and choice of variance structures in mixed models because of a substantial influence on SE of estimates.

Figure 1 provides a scatter plot of the relationships between SNP effect solutions and their SE for different levels of heritability. In all cases, SNP effect solutions were positioned along the diagonal. Furthermore, rank correlations between SNP effect estimates based on alternative heritabilities with those resulting from base heritabilities ranged from 0.99 to 1.00, demonstrating that heritability used in the analysis had only a minor influence on SNP effect estimates.

Generally, the relationship between heritability and SE of estimates was rather clear (Figure 1). For heritabilities less than base values, SE of SNP effects were

Table 1. Effect of the level of heritability (h^2) used in mixed model analyses on SNP effect estimates and their SE[1]

	Line I				Line II			
			SE of estimates				SE of estimates	
Change in sire variance from base	h^2	Regression of SNP effect estimates on estimates for the base heritability[2]	Mean	Regression on SE for base[3]	h^2	Regression of SNP effect estimates on estimates for the base heritability[4]	Mean	Regression on SE for base[5]
−50%	0.11	1.02	0.86	0.73	0.12	0.99	0.60	0.73
−40%	0.14	1.01	0.93	0.79	0.19	0.99	0.65	0.79
−30%	0.16	1.01	1.00	0.85	0.22	0.99	0.70	0.85
−20%	0.18	1.01	1.06	0.90	0.25	1.00	0.74	0.90
−10%	0.21	1.00	1.12	0.95	0.28	1.00	0.78	0.95
Base	0.23		1.17		0.31		0.82	
+10%	0.25	1.00	1.22	1.05	0.34	1.00	0.86	1.04
+20%	0.27	0.99	1.27	1.09	0.37	1.00	0.89	1.09
+30%	0.30	0.99	1.32	1.13	0.40	1.00	0.93	1.13
+40%	0.32	0.99	1.36	1.17	0.44	1.01	0.96	1.17
+50%	0.34	0.98	1.41	1.20	0.47	1.01	0.99	1.21

[1]BW measured in grams.
[2]Mean SE of estimates is 0.0001.
[3]Mean SE of estimates is 0.0011.
[4]Mean SE of estimates is 0.00002.
[5]Mean SE of estimates is 0.0004.

smaller than SE obtained for the base heritability; hence, data points were below the diagonal line. In contrast, if alternate heritabilities were greater than the base values, SE of SNP effects were greater and above the diagonal.

Relationships between differences in log-likelihood values (***dLogL***) between the full and reduced models for some alternate heritability values and the corresponding *dLogL* values for the base heritabilities are also shown in Figure 1. In all cases, there was a strong positive association between *dLogL* for alternate heritabilities, with rank correlations ranging from 0.98 to 1.0. The general pattern of the effect of heritability on the deviation of these relationships from the diagonal resembled that of the SE of SNP effects, but in the reverse direction. Regression of *dLogL* for alternate heritabilities on the corresponding values obtained from the base heritability had model R^2 of 0.99 to 1.00.

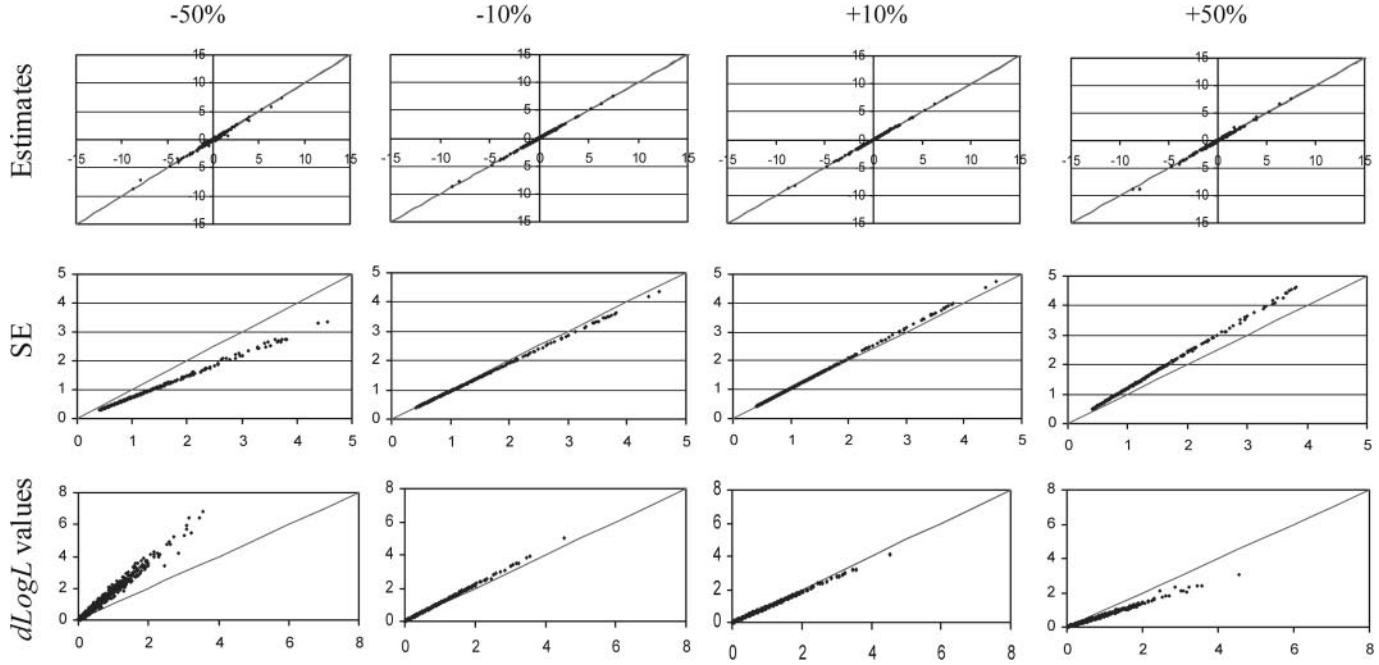

Figure 1. Relationships between results for base heritability (x-axis) and alternate heritability (y-axis) for estimates of SNP effects (first row), SE of SNP effects (second row), and differences in log-likelihood values (*dLogL*) values for the full model (third row). Plots in columns 1 to 4 are based on results when sire variances are changed by −50, −10, +10, and +50% from the base level heritability.

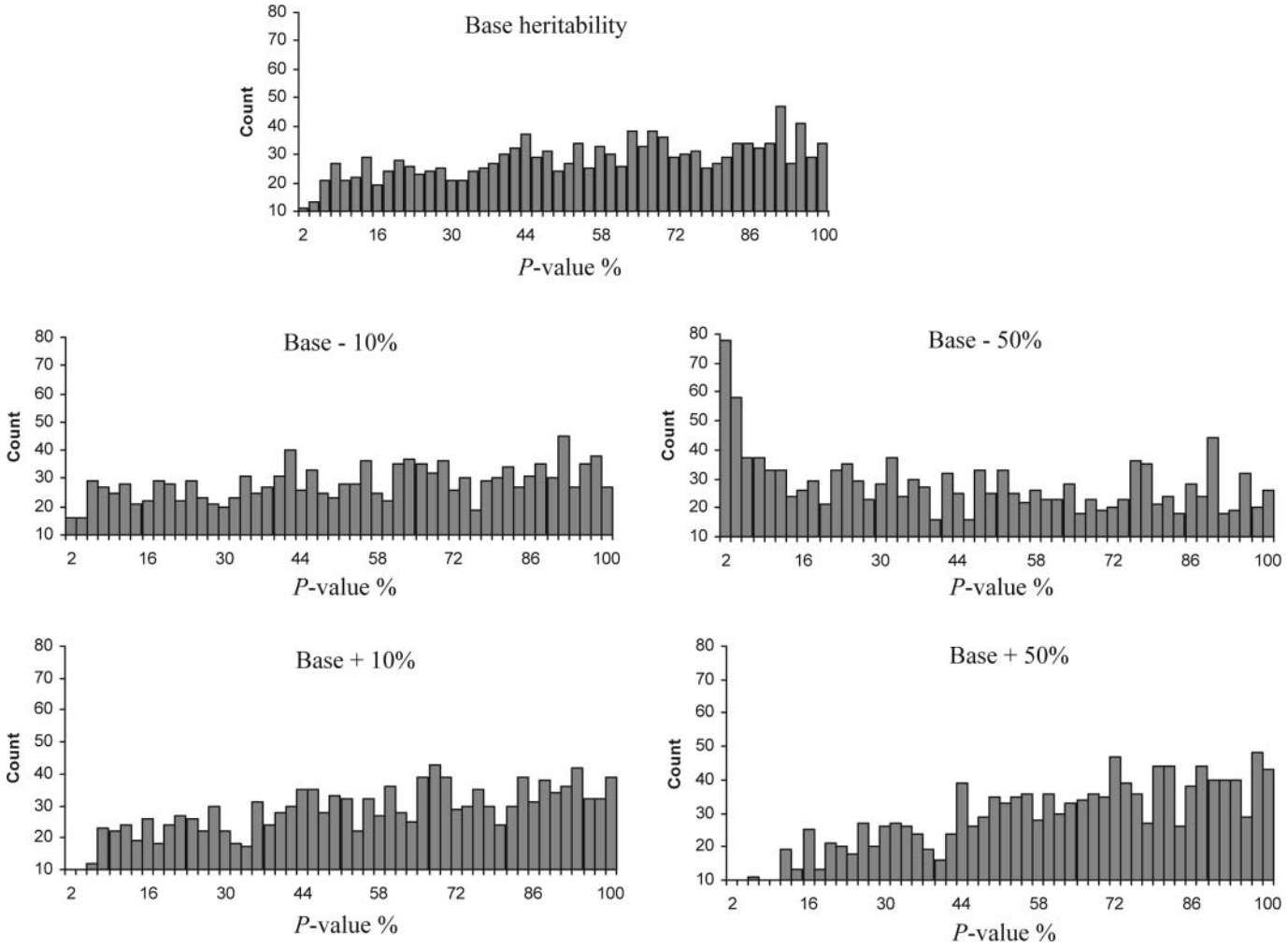

Figure 2. Frequency distribution of *P*-values for tests of SNP effects when sire variances are changed by −50 to +50% from the base.

Frequency Distributions of P-Values

The observed absence of a measurable influence of heritability on SNP effect estimates, but a substantial impact on the SE of estimates, resulted in a substantial effect on the number of SNP showing a significant association. Distributions of significance probabilities (*P*-values) for tests of SNP effects for some extreme and intermediate heritabilities are shown in Figure 2. When heritability changed from low to medium to high, the shape of the distribution of *P*-values changed from a highly skewed L-shaped curve, with a relatively high frequency of SNP showing significance, to a nearly uniform distribution, and distributions skewed to the right. When the SE of estimates increased because of high heritability, the number of SNP showing significant association decreased, such that when the greatest heritability was used, only 0.4% of the SNP had *P*-values less than 0.05, compared with only 1.7% when the base heritability was used. When the lowest heritability was used, however, approximately 9.6% of the segregating SNP had *P*-values less than 0.05. Under the null hypothesis of no association, 5% of all tests are expected to be significant at $P < 0.05$. The less than 5% significant tests for the base and greater heritabilities indicate that tests were too conservative at these levels of heritability because of overestimation of SE. Similarly, Kennedy (1991) observed changes in *F*-statistics and corresponding *P*-values when testing fixed treatment effects based on wrong variance ratios.

The practical implication of the relationship of heritability with the significance of associations between SNP genotype and performance records requires critical evaluation. The current results demonstrate that the level of heritability used in mixed model analysis influences how conservative or liberal significance tests for association can be. High heritability values lead to very conservative evaluations of significance of SNP effects, and relatively few SNP may be declared candidates for further evaluation. In contrast, tests based on low heritability may lead to a high percentage of false positives. Caution should be used to avoid both situations. For instance, according to Figure 2, the base heritabilities resulted in a *P*-value distribution that was skewed to the right, causing tests of SNP effects to be too conservative. This suggests that, although the base heritabilities were estimated by using the whole population, they may not have been the best parameter

estimates for the data used for analysis, which included data from selected sires. In fact, estimates of heritability of the trait derived from the progeny mean data used in the current study by ASReml (Gilmour et al., 2002) were 0.07 and 0.20, for lines I and II, respectively. These estimates correspond to the lower end of the alternate heritabilities used in the current analyses and would lead to tests that are less conservative than using heritability estimates from the entire population. Standard errors of estimates of heritability from these data sets were, however, relatively large because of their limited size: 0.04 and 0.13 for lines I and II.

Rescaling Results for Alternative Heritabilities

When there is uncertainty about the proper estimate of heritability to use for analysis, one possibility would be to evaluate results for different levels of heritability to identify SNP whose associations are robust to alternate levels of heritability. With the large number of SNP, traits, and lines that must be evaluated, this would, however, be computationally demanding. The strong linear relationship that is observed in Figure 1 between *dLogL* for different levels of heritabilities, however, indicates that results obtained for alternate levels can be approximated from results obtained for the base heritability without having to reanalyze the data for each level of heritability. To develop this empirical relationship, analyses could be performed for a range of heritabilities by using SNP in one or more chromosomes within each line. The *dLogL* values obtained for each SNP (i) for an alternate heritability [$dLogL_{alt}(i)$] can then be regressed on *dlogL* for that SNP for the base heritability [$dLogL_{base}(i)$] to develop rescaling values β_{alt} from the following regression model:

$$dLogL_{alt}(i) = \beta_{alt} \times dLogL_{base}(i) + e_i. \quad [1]$$

For the remaining chromosomes or groups of SNP, analyses can then be performed only for the base heritability, and $dLogL(i)$ so obtained could then be used to predict results for a different level of heritability based on $dLogL_{alt}(i) = \beta_{alt} \times dLogGL_{base}(i)$.

To evaluate the effectiveness of this procedure, *dLogL* for SNP on chromosome 1 based on each alternate level of heritability were regressed on *dLogL* obtained for the base heritability. The resulting regression coefficients are presented in Table 2. In all cases, the R^2 for the regression model was 0.99, indicating a nearly perfect fit. Slopes exceeded 1 whenever the alternate heritability was less than the base heritability, as expected based on the relationships depicted in Figure 1. Results from this procedure were validated based on 280 (line I) and 299 (line II) segregating SNP from chromosome 4. First, *P*-values for SNP on chromosome 4 were calculated by rerunning the actual analysis for the range of alternate heritabilities. The *dLogL* values obtained for the base heritability were then multiplied by the corresponding rescaling factors developed from chromosome 1 to provide rescaled *P*-values (**r*P*-values**). Correlation coefficients between r*P*- and *P*-values were 0.99 and regression coefficients between r*P*- and *P*-values were close to 1.00 (0.97 to 1.01) regardless of the level of heritability used, demonstrating the accuracy of the rescaling procedure.

Table 2. Estimates of coefficients of regression for differences in log-likelihood values of SNP effects for alternate levels of heritability on those obtained under the base level of heritability for lines I and II[1]

Change in sire variance from base, %	Line I	Line II
−50	1.78	1.88
−40	1.54	1.60
−30	1.35	1.39
−20	1.21	1.23
−10	1.09	1.10
Base		
+10	0.92	0.92
+20	0.85	0.84
+30	0.80	0.78
+40	0.75	0.73
+50	0.70	0.68

[1]Mean SE of regression coefficient estimates were 0.002.

Choice of Heritabilities

Although heritability estimated from the data included in the SNP analysis may theoretically be the most appropriate, if the data set is of limited size and represents a selected sample of the whole population, this estimate may be biased, will have a large SE, and ignoring this may bias SE and significance tests of SNP effects. This is particularly true for BW, which carries a large weight in selection decisions in broilers. To add complexity to the issue, data on SNP associations may be available from multiple lines or breeds, as was the case in our study. The interest may then be to rank SNP effects both within and across lines and to combine results across lines, after separate analysis of each line. Because of major differences between lines, heritabilities are not expected to be the same for each line. Thus, heritability values used in the analysis of each line must be chosen such that tests of SNP effects can be pooled across lines, without overemphasizing one line over another because of the use of too low an estimate of heritability for some lines and too high a heritability for other lines.

One strategy to accomplish a proper assessment of significance values for each line would be to determine and then use the heritability for each line that leads to a uniform distribution of *P*-values, thereby satisfying the null hypothesis of no significant association between SNP genotypes and phenotypes. This situation may be appropriate if a large number of SNP are tested and the number of QTL expected is to be limited or LD does not extend over large distances so that only

limited numbers of SNP are expected to have a true association. When combined across lines, true associations that are consistent across lines should still emerge as being significant with this strategy. In the current analysis, a uniform distribution of *P*-values was determined as having 50% of *P*-values above and below 0.5. This uniform distribution was reached by trial and error and was achieved at heritabilities of 0.18 and 0.25 for lines I and II, respectively. For both lines, these heritabilities lie between the base heritability estimated from the entire population and heritability estimated from the progeny mean data.

An efficient strategy to find the heritability that provides *P*-values that conform to the uniform distribution without reanalysis of data for many heritability values is to find the rescaling value β_{alt} of *dLogL* for the regression Eq. 1 that leads to a uniform distribution. In the current analysis, resulting estimates of β_{alt} were 1.41 and 1.22 for lines I and II. Thus, using these coefficients, *dLogL* values would be scaled up by 41 and 22% compared with the *dLogL* values obtained from the base heritabilities for both lines, thereby increasing the number of SNP showing significance. This is as expected from the distribution of *P*-values shown in Figure 2, which indicates that SNP effect tests resulting from the base heritability are too conservative.

If true associations are present in the data at hand, the assumption of no association that is used in the analysis above would lead to a conservative evaluation of SNP effects. To avoid this, another approach would be to create data under the null hypothesis of no association by simulating sire SNP genotypes at random, using them for analysis of associations with the actual phenotypic records, and finding the level of heritability for the simulated data that results in a uniform distribution of *P*-values. The number of SNP showing significant associations would then be as expected owing to chance, allowing the level of heritability that leads to a uniform distribution of *P*-values within a line to be derived.

One approach to generate data on the null hypothesis of no association is by random permutation (Fisher, 1935) of SNP genotypes across phenotypes. Thus, a single permuted data set was created and the heritability that resulted in a uniform distribution of *P*-values was determined by reanalysis. In agreement with the results in Figure 2, distributions of *P*-values obtained for the permuted data showed a gradual change in skewness from left to right as heritability used increased. Heritabilities of 0.21 for line I and 0.25 for line II resulted in an approximately uniform distribution. These heritabilities were only slightly greater than or similar to the corresponding values for the actual data (0.18 and 0.25, respectively).

A problem with this simple permutation approach is that resulting SNP genotypes may not be consistent with pedigree relationships. To evaluate whether this affected results, SNP genotypes were also simulated based on the existing pedigree relationships among sires. Allele frequencies for founders in a 4-generation pedigree were sampled from a uniform distribution for 1,000 SNP on 1 chromosome at 1-cM spacing. Genotypes of founders were then sampled assuming the Hardy-Weinberg equilibrium and genotypes for the remaining sires were sampled based on genotypes of their parents following Mendelian rules of inheritance and recombination. Heritabilities that gave uniform distributions of *P*-values when associating the simulated data with the actual phenotypic data were 0.21 for line I and 0.28 for line II. The fact that these heritabilities were slightly larger than the heritabilities that gave a uniform distribution of *P*-values for the actual data (0.18 and 0.25) indicates that some true effects were present in the real data.

Pooling Results Across Populations

As is the case in the current study, SNP genotypes and phenotypic data may be available from multiple lines or breeds with unique genetic parameters. Once data from each line are analyzed based on the most appropriate line-specific heritabilities, results could be combined across lines to provide an overall level of significance of SNP effects. To obtain this pooled level of significance, for each SNP j from line i, the $dLogL_{ij}$ obtained from direct analysis or rescaling (as described previously) can be summed across k lines that segregate for the SNP to obtain a pooled χ^2 statistic:

$$\chi^2 = -2\sum_{i=1}^{k} dLogL_i.$$

Under multivariate normality, the resulting statistic is expected to follow a χ^2 distribution with k df. This would be equivalent to conducting a joint analysis that allows for a different effect of the SNP and different variance components for each line, thereby accounting for potential differences in LD between lines and for effects of epistasis and genotype × environment interactions.

General Remarks

The increasing volume of high-density SNP data requires accurate and efficient data analysis procedures. When genotyped individuals are related, analyses need to account for these relationships by using proper estimates of variance components, including heritability. The current study demonstrates that the level of heritability used in mixed model analyses of SNP associations in a progeny mean design has a limited impact on estimates of SNP effects, but can substantially affect their SE and thereby the significance of SNP effects. Note that accurate *P*-values are also needed for methods to account for multiple testing, such as false discovery rate (Benjamini and Hochberg, 1995).

Ideally, Bayesian approaches would be used to account for the uncertainty of heritability estimates in association analyses. They can fully account for the uncertainty of variance estimates when inferences to the marginal posterior distribution are made by weighting all possible values of variances by their probabilities (Blasco, 2001). Such approaches are, however, computationally demanding, in particular when analyzing thousands of SNP for multiple traits and lines. In this study, computationally efficient approaches were developed to identify levels of heritability that resulted in a fair assessment of SNP effects across lines or breeds. To this end, this study demonstrated that significance results for alternative heritabilities can be approximated by rescaling results obtained from a base level of heritability, using the strong linear association between *dLogL* and heritability values. This approach substantially reduces the time and effort needed for reanalysis of data.

In this study, variance components were estimated under a model without SNP effects. In principle, variance components could be reestimated for each SNP locus, but this would be computationally demanding for high-density SNP data. In addition, for traits affected by many QTL, single SNP are not expected to explain large amounts of variation, but when the size of the data are limited, estimates of heritability may vary substantially between SNP, affecting the level of significance obtained for each SNP.

The approaches developed here also allow SNP association results to be pooled across related lines or breeds to increase power by using the heritability for each line that results in a uniform distribution of *P*-values under the null hypothesis of no association for that line. Data under the null hypothesis can be obtained by data permutation or by simulating genotypes.

LITERATURE CITED

Andreescu, C., S. Avendano, S. Brown, A. Hassen, S. J. Lamont, and J. C. M. Dekkers. 2007. Linkage disequilibrium in related breeding lines of chickens. Genetics 177:2161–2169.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B. Methodological 57:289–300.

Blasco, A. 2001. The Bayesian controversy in animal breeding. J. Anim. Sci. 79:2023–2046.

Dekkers, J. C. M., H. H. Zhao, and R. L. Fernando. 2006. Linkage disequilibrium mapping of QTL in livestock. Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil. http://www.wcgalp8.org.br/wcgalp8/articles/paper/21_896-1316.pdf Accessed Jan. 13, 2002.

Fisher, R. A. 1935. The Design of Experiments. 3rd ed. Oliver and Boyd Ltd., London, UK.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2002. ASReml User Guide Release 1.0, VSN International Ltd., Hemel Hempstead, UK.

Kackar, R. N., and D. A. Harville. 1981. Unbiasedness of two stage estimation and prediction procedures for mixed linear models. Commun. Stat. 10:1249–1261.

Kennedy, B. W. 1991. The legacy of C. R. Henderson: The unfinished legacy. J. Dairy Sci. 74:4067–4081.

Kennedy, B. W., M. Quinton, and J. A. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci. 70:2000–2012.

Littell, R. C., P. R. Henry, and C. B. Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. J. Anim. Sci. 76:1216–1231.

Littell, R. C., J. Pendergast, and R. Natarajan. 2000. Modeling covariance structure in the analysis of repeated measures data. Stat. Med. 19:1793–1819.

Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28:161–176.

Meuwissen, T. H. E., B. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Searle, S. R. 1987. Linear Models for Unbalanced Data. John Wiley and Sons, New York, NY.

Sorensen, D. A., and B. W. Kennedy. 1986. Analysis of selection experiments using, mixed model methodology. J. Anim. Sci. 63:245–258.

Wong, G. K., B. Liu, J. Wang, Y. Zhang, X. Yang, Z. Zhang, Q. Meng, J. Zhou, D. Li, J. Zhang, P. Ni, S. Li, L. Ran, H. Li, J. Zhang, R. Li, S. Li, H. Zheng, W. Lin, G. Li, X. Wang, W. Zhao, J. Li, C. Ye, M. Dai, J. Ruan, Y. Zhou, Y. Li, X. He, Y. Zhang, J. Wang, X. Huang, W. Tong, J. Chen, J. Ye, C. Chen, N. Wei, G. Li, L. Dong, F. Lan, Y. Sun, Z. Zhang, Z. Yang, Y. Yu, Y. Huang, D. He, Y. Xi, D. Wei, Q. Qi, W. Li, J. Shi, M. Wang, F. Xie, J. Wang, X. Zhang, P. Wang, Y. Zhao, N. Li, N. Yang, W. Dong, S. Hu, C. Zeng, W. Zheng, B. Hao, L. W. Hillier, S. P. Yang, W. C. Warren, R. K. Wilson, M. Brandström, H. Ellegren, R. P. Crooijmans, J. J. van der Poel, H. Bovenhuis, M. A. Groenen, I. Ovcharenko, L. Gordon, L. Stubbs, S. Lucas, T. Glavina, A. Aerts, P. Kaiser, L. Rothwell, J. R. Young, S. Rogers, B. A. Walker, A. van Hateren, J. Kaufman, N. Bumstead, S. J. Lamont, H. Zhou, P. M. Hocking, D. Morrice, D. J. de Koning, A. Law, N. Bartley, D. W. Burt, H. Hunt, H. H. Cheng, U. Gunnarsson, P. Wahlberg, L. Andersson, E. Kindlund, M. T. Tammi, B. Andersson, C. Webber, C. P. Ponting, I. M. Overton, P. E. Boardman, H. Tang, S. J. Hubbard, S. A. Wilson, J. Yu, J. Wang, H. Yang, and International Chicken Polymorphism Map Consortium. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432:717–722.

Zhang, Z. 2005. BLUP with SAS . J. Anim. Sci. 83(Suppl. 1.):146. (Abstr.)